AD-A256 953

(12)

SECURIT

DOCUMENTATION PAGE

1a. REPC — UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHO

2b. DECLASSIFICATION/DOWNGRA — SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for Public Release:
Distribution Unlimited

DTIC ELECTE OCT 16 1992 S D

4. PERFORMING ORGANIZATION REPORT NUMBER

5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION
Department of Statistics

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (City, State and ZIP Code)
University of North Carolina
Chapel Hill, North Carolina

7b. ADDRESS (City, State and ZIP Code)

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
Office of Naval Research

8b. OFFICE SYMBOL (If applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
N00014-89-J-1175

8c. ADDRESS (City, State and ZIP Code)
Statistics & Probability Program
Arlington, VA 22217

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.

11. TITLE (Include Security Classification)
A Computational Evaluation of New Detection Algorithms

12. PERSONAL AUTHOR(S)
C.R. Baker, M.R. Frey and C.E. Persons

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
TECHNICAL	FROM _____ TO _____	September 1990	20

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	NonGaussian signal detection, Likelihood ratios, Passive sonar, Sonar signal processing

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This paper summarizes results of a computational study of two new signal detection algorithms. The new algorithms have the potential for significantly improving existing detection methods when the signal-plus-noise process is broadband (stationary or non-stationary), especially when it is nonGaussian. They require no assumptions on the statistical properties of the signal-plus-noise process; instead, they require that the drift function of a diffusion be known or estimated. When this function is known, the new discrete-time algorithms are approximations to the likelihood ratio for the continuous-time data under some reasonable assumptions on the data characteristics. These assumptions include that of Gaussian noise, although the computational results indicate that good performance can be obtained when the noise is not Gaussian. The study included comparisons with several reference algorithms, using both simulated and passive sonar data. The new methods gave superior performance despite the use of a very rudimentary procedure

CONTINUED ON BACK

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
C.R. Baker	(919) 962-2189	

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

19. ABSTRACT CONTINUED:

for estimating the drift function. Further improvements are expected when the estimation procedure is optimized. One of the new algorithms is fully adaptive to the signal-plus-noise process; it relative performance can be expected to further improve when used with longer observation times.

U.S. Navy Journal of Underwater Acoustics Volume 41, No. 3 July 1991

A COMPUTATIONAL EVALUATION OF NEW DETECTION ALGORITHMS

Charles R. Baker

Department of Statistics
University of North Carolina
Chapel Hill, NC 27599-3260

Michael R. Frey

National Institute of Standards and Technology
Admin. Bldg. A334
Gaithersburg, MD 20899

Charles E. Persons

Processing Applications Branch
Naval Ocean Systems Center
San Diego, CA 92152

(Received 11 September 1990)

This paper summarizes results of a computational study of two new signal detection algorithms. The new algorithms have the potential for significantly improving existing detection methods when the signal-pulse-noise process is broadband (stationary or nonstationary), especially when it is non-Gaussian. They require no assumptions on the statistical properties of the signal-plus-noise process; instead, they require that the drift function of a diffusion be known or estimated. When this function is known, the new discrete-time algorithms are approximations to the likelihood ratio for the continuous-time data under some reasonable assumptions on the data characteristics. These assumptions include that of Gaussian noise, although the computational results indicate that good performance can be obtained when the noise is not Gaussian. The study included comparisons with several reference algorithms, using both simulated and passive sonar data. The new methods gave superior performance despite the use of a very rudimentary procedure for estimating the drift function. Further improvements are expected when the estimation procedure is optimized. One of the new algorithms is fully adaptive to the signal-plus-noise process; its relative performance can be expected to further improve when used with longer observation times.

INTRODUCTION

Recent improvements in quieting of noise radiated by submarines have changed the nature of the passive sonar detection and classification problem. Narrowband filtering followed by a power detector works well when the emitted noise consists largely of signals with line spectra. However, if the emitted noise is primarily broadband, then such a simple algorithm is no longer effective in low signal-to-noise ratio (SNR) applications. These problems require detector and classification algorithms that are effective for a broadband, nonstationary, stochastic signal imbedded in additive noise. For the passive sonar application in a quiet ocean, the signal consists of the noise emitted by the submarine, while the additive noise consists of receiver noise and ambient ocean noise. Frequently, the additive noise process is Gaussian or near-Gaussian.

This paper summarizes the results of a computational evaluation of two new discrete-time detection algorithms that may contribute to the solution of the new-era passive sonar detection problem. These are likelihood-ratio-based algorithms under the assumption of Gaussian noise. However, in contrast to the usual requirements on likelihood-ratio-based algorithms, their optimality is not based on knowledge of the statistics of the signal-plus-noise (S + N) process. Instead, optimality is based on knowledge of the drift function of a diffusion. Their implementation requires knowledge or estimation of this drift function and knowledge or estimation of the noise covariance matrix and mean vector. In practice, these parameters are typically estimated from data, and this is the procedure used in the study reported here.

Although the work reported here considered only detection, the new algorithms have obvious potential for classification. The signal component of the S + N process is represented by a filtered diffusion drift in the equations leading to the algorithms, and different target classes would correspond to different diffusion drift functions.

The study included comparisons with appropriate reference algorithms. The evaluations resulted in the new algorithms clearly outperforming comparable algorithms for detection of a broadband signal at low values of false alarm probability (P_{FA}). This was despite the fact that the work did not include optimization of the method used to estimate the diffusion drift function. It is speculated that the algorithms' already excellent performance can be further improved with optimization of the estimation procedure.

The derivation of the two new algorithms is partially contained in Refs. 1 and 2. Reference 3 contains a detailed discussion, including a derivation. They are optimum (approximations to a log-likelihood ratio) for detecting stochastic signals in Gaussian noise under some mild assumptions on the nature of noise and the S + N processes.[2,3] These assumptions include: mean-square continuity of the continuous-time noise process from which the noise vector is obtained by sampling; zero energy in the noise process at time zero (beginning of the observation period); spectral multiplicity of one (in the sense of Cramér and Hida[1,2,3]) for the continuous-time process. The first of these three assumption is typically satisfied; the third is approximately true in a mean-square sense;[2,3] and the second can be fitnessed (when not satisfied) by assuming that the first actual sample occurs at the second sampling time.

The algorithms were evaluated by using simulated Gaussian data and, more extensively, using passive sonar data obtained from the output of a single hydrophone. The recording consisted of a segment of noise (N), followed by a segment of a signal-plus-noise (S + N), followed by another segment of noise.

Five statistical tests for univariate normality were conducted on both the noise and the signal-plus-noise data. In general, neither the noise nor the signal-plus-noise could be clearly accepted as Gaussian.

The tendency toward Gaussian varied, depending on the frequency range being investigated. The non-Gaussian nature of the data is illustrated by the relatively poor performance of the optimum (likelihood ratio) detection algorithm under the hypothesis of Gaussian data (N and S + N), as shown below.

One of the new algorithms is totally adaptive to the signal-plus-noise process; implementation of the other requires a "training" ensemble of signal-plus-noise data or prior knowledge of a time-varying function representing a diffusion drift function. Both require knowledge or estimation of the noise covariance matrix and mean vector. Comparisons were made with performance of reference algorithms requiring comparable knowledge about the signal-plus-noise process. The algorithm that is adaptive to the S + N process will henceforth be referred to as "adaptive." Although it is not fully adaptive, the parameters that are required for its implementation depend solely on the noise, and are typically (for the problems of interest here) much easier to obtain in reliable form than significant parameters of the S + N process.

For the adaptive algorithm, termed Version I, comparisons are made with an algorithm that computes the squared norm of the output of a noise whitener (which is implemented by following a noise-whitener with square-law device and then by integration), denoted WEN. The WEN requires prior knowledge of the noise covariance matrix and mean vector, the same information required by the Version I algorithm. Another reference algorithm was a simple energy detector, EN, which omits the noise whitener.

For the nonadaptive algorithm, denoted Version II, comparisons were made with the classical log-likelihood-ratio detection algorithm (denoted GvG) when the data (noise and S + N) are Gaussian, and with the best quadratic-plus-linear detector in Gaussian noise based on the deflection criterion (denoted as DFL). These algorithms require the same type of knowledge for their implementation as does the Version II algorithm, albeit partially in different form. All three require knowledge of the noise mean vector and covariance matrix. In addition, the two reference algorithms require knowledge of the S + N mean vector and covariance matrix. The Version II algorithm also requires knowledge of a two-variable function determined by the S + N process: rather than a covariance matrix, this is a diffusion drift function. Since none of these parameters are likely to be known in applications, implementation of all three algorithms typically requires an ensemble of noise data and an ensemble of S + N data from which to estimate the parameters.

Based on the assumptions, the continuous-time S + N process can be represented as a filtered diffusion. The diffusion has the general form $Z(t) = \int_0^t \sigma(s, Z(s))ds + W(t)$, where the function σ is the drift function of the diffusion and W is the standard Wiener process.[2,3] Effective estimation of this function is the major problem in realizing the potential of the new algorithms. To implement Version II, the drift function must be estimated (or, ideally, known). Version I estimates this function from the observation vector, under the assumption that it is time-invariant. Polynomial regression was used to estimate the drift function in the work reported here. *However*, for polynomials of order greater than one, a polynomial drift function does not satisfy the assumptions under which the algorithms were derived when reasonable physical constraints are imposed. Thus, the results given here should be considered only as *lower bounds* on their achievable performance.

With a large training ensemble, each of the two algorithms clearly outperformed its competitors on the sonar data at low values of P_{FA}. Their relative performance tailed off at high values of P_{FA}. It is speculated that this is due to poor estimation of the drift function. Additional performance gains should be possible with more sophisticated estimation procedures.

With small training ensembles, the adaptive algorithm outperformed all others at low values of P_{FA}. This is one of the most striking and encouraging (for eventual applications) results of the study.

The algorithms considered in this study that require prior knowledge of the S + N process are Version II, DFL, and GvG. In the study, this was obtained from an ensemble of training data. Such an ensemble, almost perfectly matched to the evaluation data, will rarely, if ever, be available in applications. Alternatively, the required parameters (mean vectors, covariance matrices, and diffusion drift function) can be obtained from a good mathematical model of the S + N process. Again, this is not likely to be available in many important applications. For such applications, the relative performance of GvG, DFL, and Version II can only be regarded as benchmarks to which more implementable algorithms can be compared. A possible exception is the Version II algorithm using a time-invariant drift function, whose performance actually compared rather well with that of Version II when time-varying drift was used. This implementation of Version II may be a reasonable goal for some applications. However, as will be discussed, it is speculated that the adaptive Version I may have performance comparable to that of Version II with time-invariant drift when long observation times are available.

Thus, a reasonable hypothesis is that the adaptive Version I is the algorithm having the most potential for applications, and that its performance for long observation times is likely to be superior to that of all the reference algorithms evaluated here, even when those algorithms have large S + N training ensembles available. Of course, this must be qualified as resting largely on the assumption that the results here are indicative of performance in more general applications.

These computational results, although very encouraging, should not be regarded in any sense as definitive. Their principal contribution is to give a numerical confirmation, based on actual sonar data, of the theoretical potential of the new algorithms. The fact that the algorithms performed so well with very little attention given to optimizing their performance is especially encouraging, as is the relative performance of the adaptive version in the face of an extremely short observation time. Since the new algorithms require no assumptions on stationarity or on the signal being composed of a set of narrowband components, they have obvious potential for applications to some of the Navy's most pressing detection and classification problems. A long-term comprehensive program is needed to fully develop the algorithms for sonar applications. Of course, many more data sets should be used for evaluations and comparisons. Beyond this, a mixture of computational and theoretical research is needed to optimize performance.

DATA ENSEMBLES

This section describes the data ensembles used in the detection studies. The studies, using experimental data, were carried out on passive sonar data. These data are a time series of real numbers that had been digitized from an analog tape recording of a single hydrophone. The recording was made when the broadband-radiating target platform (at an unknown depth) passed by the omnidirectional hydrophone in a deep-ocean basin. The analog recording was made with instrumentation that preserved frequency stability and provided a bandwidth well in excess of target frequencies of interest. A single-channel analog-to-digital converter provided the time series data, which were stored on a nine-track tape.

A lofargram obtained from an array containing the above-mentioned hydrophone was obtained for this same event. From inspection of the lofargram, noise (N) and signal-plus-noise (S + N) data segments were identified. The noise used in the study was obtained from a data segment of approximately 4.07 minutes duration, immediately preceding the S + N data segment. The latter was

of the same time duration as the noise data. The lofargram and the hydrophone recording were furnished by the Naval Ocean Systems Center (NOSC).

For the large training ensembles, the available NOSC data were divided into four ensembles, each ensemble consisting of 5000 vectors of length 100. Since the sampling rate was 4096 samples/second, each vector represented continuous-time data for a length of 100/4096 second.

The four ensembles consisted of two for training and two for evaluation, and were formed as follows. For the noise training ensemble, a segment containing $5000 \times 100 \times 2 = 10^6$ sample values was selected. The first 100 sample values were selected for the training ensemble, the next 100 for the evaluation ensemble, the third 100 for the training ensemble, and so on. Thus, alternating 100-component segments were selected for the training ensemble, alternated with 100-component segments selected for the evaluation ensemble. A similar procedure was followed in forming the training and evaluation ensembles for the signal-plus-noise.

For the small-training-ensemble evaluations, the evaluation ensembles were those defined above (N and S + N ensembles, each consisting of 5000 sample vectors). However, the training ensemble for the noise was formed by taking only the first 200 vectors of the 5000-vector training ensemble used in the large training ensemble. Similarly, the S + N training ensemble was formed by taking only the first 200 vectors of the 5000-vector training ensemble used in the large-training-ensemble evaluations described above.

Figures 1 through 5 show the results for unfiltered data. The algorithms evaluated here were also evaluated using low-pass filtered data. Figure 6 shows some of these results. (For convenience, all figures are grouped in the PERFORMANCE RESULTS section).

In addition to the ensembles formed from experimental data, two ensembles (one N, the other S + N) of 100-component vectors were generated by computer simulation. The N ensemble was from the Wiener process with mean zero, variance 1/100 (the sampling interval). The S + N ensemble was from a diffusion with drift function f, $f(x) = -25x$. The algorithms were also evaluated for detection performance on this data set.

DEFINITIONS AND DETECTION ALGORITHMS

This section contains a definition of basic quantities used to define the detection algorithms and a definition of the test statistic Λ formed by each algorithm.

Basic Quantities

R_N: Noise covariance matrix

R_{S+N}: S + N covariance matrix

m_N: Noise mean vector

m_{S+N}: S + N mean vector

Δ: Sampling interval

$\underline{\delta X}$: Vector of increments of a process X, obtained by sampling at interval Δ. $[\delta X](k) = X([k + 1]\Delta) - X(k\Delta)$.

$\underline{W}$: Random vector obtained by sampling the standard Wiener process W(t). $\underline{\delta W}$ has components and are i.i.d. (independent and identically distributed), normal, and with variance Δ. The mean of $\underline{\delta W}$ in the experimental data was found to be non-negligible and was subtracted out. That is, the noise mean was estimated from the training data and this mean was subtracted from the evaluation data, so that $\underline{N}$ and $\underline{\delta W}$ were treated as if having zero mean.

F: Lower triangular matrix satisfying $R_N = \Delta FF^*$ and $\underline{N} = F\underline{\delta W}$, where * denotes transpose. $\underline{N} = F\underline{\delta W}$ is a discrete-time representation of the noise process, which is assumed to have the continuous-time representation $N(t) = \int_0^t F(t,s)dW(s)$. The F in the latter representation is a function on $[0, T] \times [0, T]$, where T is the time duration of the observed waveform.

σ: Drift function of the diffusion Z assumed (for implementation of the new algorithms) to give the S + N process: $\underline{S + N} = F\underline{\delta Z}$, where F is defined as above, and $(\delta Z)(k) = \Delta\sigma(k, Z_k) + \delta W(k)$. $\underline{\delta Z}$ is a discrete-time representation of the differential of a diffusion process Z having the representation $Z(t) = \int_0^t \sigma(s, Z(s))ds + W(t)$, where σ is the drift function of the diffusion.

L: Summation matrix; $L(i, j) = 1 \quad i \geq j$
$= 0 \quad i < j.$

See Ref. 3 for a discussion of how diffusion processes arise in this application and Ref. 4 for a general discussion of such processes. The Wiener process in the representation for the diffusion is not the same Wiener process as that in the representation of the noise; see Ref. 3 for a discussion.

Detection Algorithms

Each detector forms a test statistic Λ having the value $\Lambda(x)$ when x is the observed vector. The decision is to decide "signal present" if $\Lambda(x)$ exceeds a threshold, decide "noise only" if it does not. For a given detector, the value of the threshold depends on the false alarm probability P_{FA}.

1. Version I and Version II Algorithms (V.I and V.II), for k-dimensional data vectors:[3]

$$\Lambda(x) = \sum_{j=1}^{k-1} \sigma[j, (LF^{-1}x)_j][F^{-1}x]_{j+1} - \frac{\Delta}{2}\sum_{j=1}^{k} \sigma^2[j, (LF^{-1}x)_j].$$

In Version I, σ is estimated from the observed vector x and is time-invariant ($\sigma(j, y) = \sigma(i, y)$ for all i, j, y).

In Version II, σ is estimated from a training ensemble of S + N vectors, is permitted to be time-varying, and is inserted into the algorithm prior to the observation of the received waveform.

2. Gauss-vs-Gauss log-likeihood ratio (GvG; see, e.g., Ref. 5):

$$\Lambda(x) = (x - m_N)^* R_N^{-1}(x - m_N)^* - (x - m_N)^* R_{S+N}^{-1}(x - m_N)^* + 2(x - m_N)^* R_{S+N}^{-1}(m_{S+N} - m_N).$$

3. Deflection Criterion Algorithm (DFL; Refs. 6, 7):

$$\Lambda(x) = (x - m_N)^* W(x - m_N)^* + (x - m_N)^* h$$

where $W = R_N^{-1}(R_{S+N} - R_N)R_N^{-1}$ and $h = 2R_N^{-1}(m_{S+N} - m_N)$.

4. Noise Whitener - Energy Detector (WEN): $\Lambda(x) = x^* R_N^{-1} x$.

5. Energy Detector (EN): $\Lambda(x) = x^* x$.

As can bee seen, the simple energy detector requires no prior knowledge of the data properties. It is perhaps closest to the standard lofargram when the input consists of broadband data. That is, the lofargram is presumably constructed by computing the energy output from a large number of contiguous narrowband filters. With broadband data having energy reasonably uniformly distributed across all frequencies of interest, the narrowband filtering would serve no useful purpose. For some of the evaluations, the EN algorithm is preceded by narrowband filtering.

The performance of the simple energy detector can be expected to lower-bound that of all the other detectors. The two detectors WEN and V.I should give the next lowest performance; they require knowledge only of the noise covariance matrix and mean vector.

The remaining three detectors, GvG, DFL, and V.II, all require knowledge of the S + N process as well as knowledge of the noise covariance matrix and mean vector. GvG and DFL require knowledge of the S + N covariance matrix and mean vector. V.II requires knowledge of the (assumed) drift function generating the diffusion which, when filtered by F, gives the S + N process.

In the studies summarized here, the σ appearing in the definition of Λ for V.I and V.II was modeled as a low-order polynomial. For the large training ensembles, the maximum order investigated was of order 8 for the V.II algorithm. This was obtained by regression on 5000 sample values. For the V.I detector, which had only 100 sample values with which to estimate σ, the maximum order of polynomial investigated was 3. The short observation time was presumably a substantial disadvantage for V.I.

For the small training ensembles, the V.II, Gauss-vs-Gauss, and Deflection algorithms were given only 200 × 100 = 20,000 sample values of the S + N data from which to estimate the parameters, depending on S + N. Together with V.I and the WEN, they were also given only 200 × 100 = 20,000 sample values of the noise from which to estimate the noise covariance matrix and mean vector. For this evaluation, the V.I algorithm outperformed all others at low values of P_{FA}. This is one of the most significant and promising aspects of the study.

ESTIMATION OF THE DIFFUSION DRIFT FUNCTION

For the two new algorithms, V.I and V.II, it is assumed that the S + N process has the form

$$\underline{Y} = F\underline{\delta Z}, \tag{1}$$

where $R_N = \Delta FF^*$, F is lower triangular, and

$$(\delta Z)(k) = \Delta\sigma(k, Z[k]) + W([k + 1]\Delta) - W(k\Delta). \tag{2}$$

$\underline{W}$ is a sampled Wiener process, so that defining $\underline{\delta W}$ by $(\delta W)(k) = W([k + 1]\Delta) - W(d\Delta)$, $\underline{\delta W}$ has independent and identically-distributed components, each Gaussian with zero mean and variance of Δ (the sampling interval).

With this model, the unknown σ was estimated by modeling it as a polynomial of various orders and the coefficients estimated by using multiple linear regression.[8] Performance using various orders was a topic of investigation. Thus, for a *p*th order polynomial, the model was

$$(\delta Z)(k) = \Delta \sum_{i=0}^{p} \sigma_{ki} Z_k^i + (\delta W)(k). \tag{3}$$

The unknowns then consist of the coefficients $\{\sigma_{ki}, i \leq p, k \leq 100\}$. They were estimated by standard polynomial regression. For the Version I (adaptive), $\sigma_{ki} = \sigma_i$ for all k, each i: the drift function σ is time-invariant.

However, a polynomial of order greater than one does not satisfy the assumptions required for the existence of the continuous-time likelihood ratio if one makes reasonable physical assumptions on the continuous-time version of the process Z (the stochastic differential equation represented by the diffusion may not have a solution; see Ref. 4. Polynomials were used primarily because of their ease of implementation. The performance of the algorithms using polynomials can be expected (for strongly non-Gaussian data) to be worse than performance using more appropriate drift function models.

Although the constraints of the study precluded a serious investigation of alternative methods of estimating the drift function, a modest deviation from ordinary polynomial regression was effected by using weights. The original motivation for this was to compensate for the inappropriate nature of polynomials of order greater than one as a model for the drift function. The procedure can be summarized as follows. Suppose that σ was represented by a *p*th-order polynomial so that for the *j*th sampling time $\sigma(j, x) = \sigma_{j,0} + \sigma_{j,1}x + \ldots + \sigma_{j,p}x^p$. First, estimate the unknown coefficients σ_{ji} using standard polynomial regression.[8] Then, multiply the coefficients $\{\hat{\sigma}_{ji}, i = 2, \ldots, p\}$ of the nonlinear terms by selected weights and use the resulting values in the implementation. Various weights are used; in some schemes, the weights vary with the coefficients; in others, the weights are the same. However, the investigation is rather limited, and the use of different weights for the coefficients of different nonlinear terms does not give appreciably better results than those obtained by using a constant weight on all coefficients of nonlinear terms. This procedure is termed modified regression; the results reported here are for constant weights.

The weights used in the study ranged from .5 to 3.33. The value of the weight used to obtain the curves presented here was either 1.6 or 2. Weights of more than one increase the absolute values of the coefficients of the nonlinear terms, as compared to unweighted coefficients. In the case of the V.I (adaptive) algorithm, the best results were obtained by using a second-order polynomial with these weights. This can be interpreted as being a consequence of two factors: the limited amount of data from which the estimation was made, and the non-Gaussian nature of the S + N process. The first factor would tend to make the use of higher-order polynomials unsatisfactory, since this would increase the number of unknowns to be estimated from the data. The non-Gaussian nature of the data, however, would lead to a need for emphasizing nonlinear effects in the drift function, and this is achieved by increasing the absolute value of the coefficient for the second-order term. It must be noted, however, that even for simulated Gaussian data the V.I algorithm performed best when using a second-order polynomial and modified regression, rather than the first-order polynomial that theory would indicate. This can be attributed to the small number of data samples available for the estimation. In the case of V.II, which had a much larger sample for making the estimation, a first-order polynomial for the drift function gave best performance on the simulated Gaussian data; a polynomial of order 7 gave best performance on the sonar data.

There are many possible methods of implementing the estimation of the drift function. The choice of polynomial regression for this study was due primarily to its ease of implementation, along with some speculation that the sonar data might be sufficiently near to Gaussian that nonlinear polynomial terms would be of secondary importance. The latter was not borne out by the results of the study. The data not only deviated from Gaussian on the basis of statistical testing, but the detection results for the Gauss-vs-Gauss log-likelihood ratio were markedly inferior to other algorithms and of a nature such that the nonlinear properties are significant. Polynomials are therefore not suitable, based on theoretical considerations; it is very encouraging that the results using them were so good. More appropriate models and methods, satisfying the conditions for existence of the solution to a diffusion stochastic differential equation and providing the nonlinear drift function needed to model a strongly non-Gaussian diffusion, should be the subject of extensive further investigations on the algorithms.

SUMMARY OF RESULTS

Presentation

Results are shown in Fig. 1 through 7. Figure 1 is for simulated data; the remaining figures are for the passive sonar data previously described. We summarize here the main results for the sonar data.

Large Training Ensembles

With a large training ensemble (5000 training vectors), the V.II algorithms with σ assumed to be a polynomial of order 7 gave the best performance of all algorithms at low P_{FA} values. The deflection detector (DFL) was second, followed (in order of performance) by V.I, GvG, WEN, and EN. The relatively poor performance of GvG is a striking indication of the non-Gaussian nature of the data and the sensitivity of this algorithm to the assumption of normality.

It is emphasized that the V.I algorithm relies totally on the observed waveform to make its estimate of σ. For the observation vectors of this study, V.I. had 100 sample values with which to work. The V.II, DFL, and GvG detection algorithms all had $5000 \times 100 = 500{,}000$ points from which to estimate their signal-dependent algorithm parameters. Thus, the relative performance of the V.I algorithm should improve markedly when used with longer observation times.

Small Training Ensembles

Two hundred noise training vectors and 200 signal-plus-noise training vectors were used in this evaluation. Five thousand vectors, the same sets used in the evaluations for large training ensembles, were used for the evaluation ensembles.

The most striking aspect of these results is the relative performance of the Version I algorithm. It outperformed all other algorithms, even Version II, at low values of P_{FA}. Evidently, the same factor that works against V.I with a large training ensemble works in its favor with a small training ensemble: lack of dependence on prior knowledge of the signal-plus-noise properties. Since the V.I algorithm is by far the most reasonable algorithm to implement (among all algorithms giving good performance at low values of P_{FA}), this result is highly encouraging for applications.

Time-Invariant Version II

Version II permits a time-varying drift function. Averaging the estimated drift function over time gives a time-invariant function, and the performance of V.II using this for the drift function can be expected to be comparable to that of V.I for long observation times. The results for the time-averaged V.II using a second-order polynomial for σ were very near those of the V.II with a time varying σ and the best-performing implementation (seventh order polynomial). Thus, for long observation times, one may speculate that the V.I algorithm with second-order drift polynomial will have relative performance compared to that of GvG and DLF, which is reasonably near that of the best polynomial implementation of V.II using a time-varying drift function.

RANDOMNESS PROPERTIES OF THE TEST STATISTICS

In considering the elements necessary to have a valid evaluation of detection performance, the following comments are relevant. For a given set of test statistic output values and a fixed threshold T, define a corresponding 0-1 value for each statistic output, depending on whether or not the test statistic output exceeds T. The resulting set of random variables should have a binomial distribution. Since the set of threshold values varies considerably as a receiver operating curve (ROC) is constructed, this will typically require that the set of test statistic output be random (independent and identically distributed, or i.i.d.).

Tests for randomness were conducted on four sets of detector outputs. Two sets consisted of test statistic values for the simulated noise and S + N data used to form Fig. 1. The other two sets were formed from the noise and S + N evaluation sonar data. Four tests of randomness were applied to each set.

For the simulated data, both sets of test statistics were accepted as random by all four tests at a significance level of .05. Each of the two sets formed from the sonar data were rejected as being random by three of the four randomness tests.

Tests for univariate randomness of the sonar data were also conducted. These tests indicated that data points separated by 24 samples could be accepted as i.i.d., provided that the total sample (including omitted samples) did not exceed 17,000 data points. For each of the two sets of test statistics, the vectors used to form the test statistics were separated by 100 of the original data samples. Thus, it seems most likely that the test statistics' outputs were independent but not identically distributed.

As discussed in Ref. 9, a failure to pass randomness tests can give insight into the nature of the data properties. However, in the present case this is not easy to see, since the test statistic for which the tests were made was the Version I algorithm using second-order polynomial regression to estimate σ. This test statistic involves linear, quadratic, cubic, and quartic operations on the original data. Even for the simplest algorithm, VI.o1 (V.I using first-order polynomial regression), the test statistic involves linear and quadratic operations, and then the difference of such operations. Thus, no general statement on the reasons for the failures seems possible, except that it is most likely not due to lack of independence.

These facts do not seriously detract from the study. As noted above, statistical tests on the data indicate that the set of test statistic outputs are independent, so that the probably cause of failing the randomness tests is lack of being identically distributed. If, however, the data are representative, this is just a consequence of the physical world. It means that data gathered over shorter observation periods are probably needed to obtain test statistic outputs that would be accepted as identically distributed by tests for randomness. Over the observation time used to take the data used in this study, the results will give an estimate of relative performance of the several algorithms. It is this relative performance, rather than individual quantitative estimates, that is of most interest here.

PERFORMANCE RESULTS

The following figures show performance of the algorithms. Some preliminary comments are appropriate.

First, it is considered that P_{FA} values of .02 and lower are of major interest.

The new algorithms were evaluated by using various orders of polynomial or modified polynomial regression to estimate the drift function. The order of the polynomial is indicated in the designation of the algorithm. For example, V.IIo1 is the Version II algorithm using first-order polynomial regression to estimate the drift. Modified polynomial regression is indicated by an additional asterisk or other identifier. These are defined in the comments immediately preceding the figure where the performance is given.

All figures except Fig. 3 show results obtained with large training ensembles of 5,000 vectors. Figure 1 is for simulated data; Figs. 2 through 7 are for the sonar data.

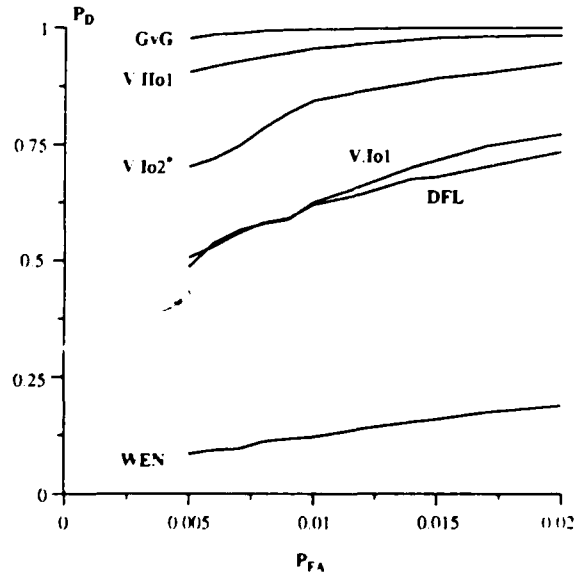

Fig. 1 — (U) Simulated data

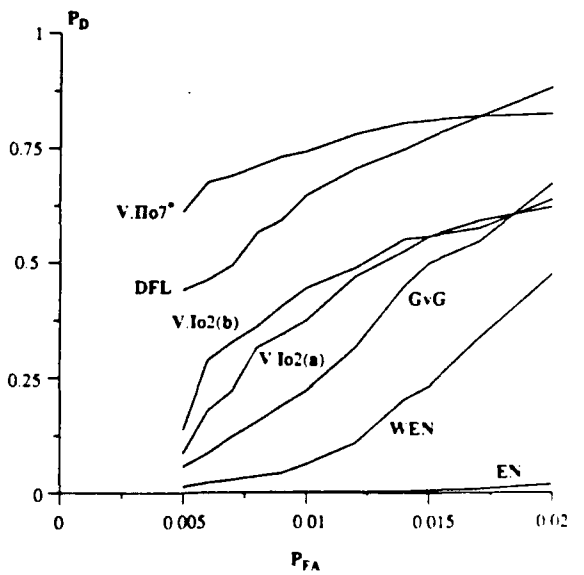

Fig 2 — Large training ensembles

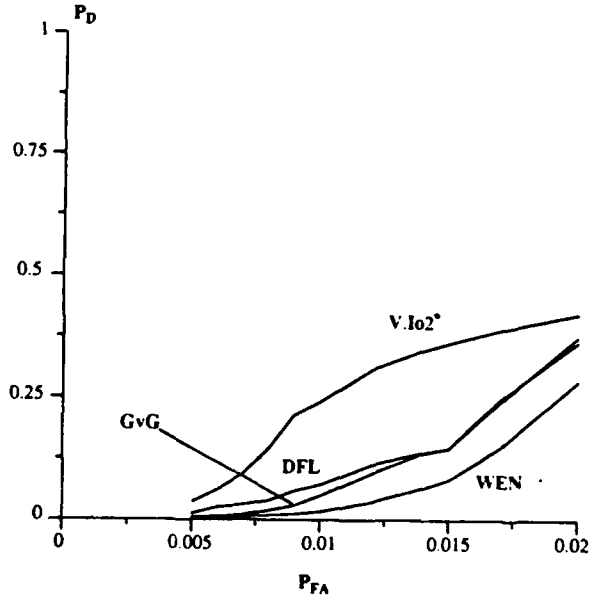

Fig. 3 — (U) Small training ensembles

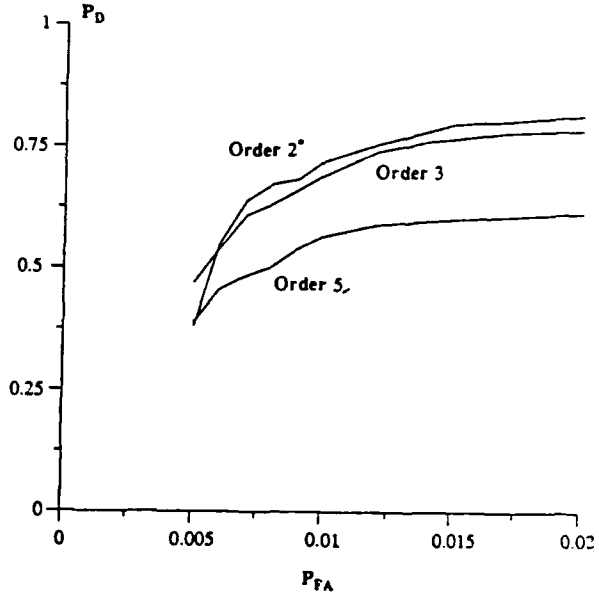

Fig. 4 —Time-averaged drift, Version II

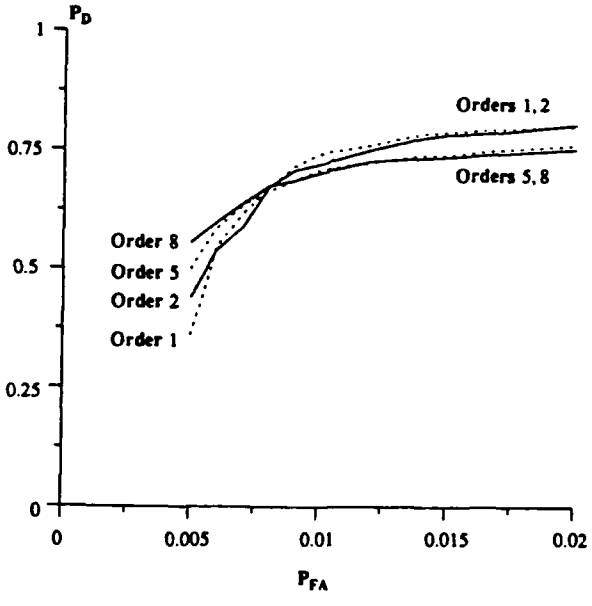

Fig. 5 — (U) Version II performance, various orders of polynomial drift

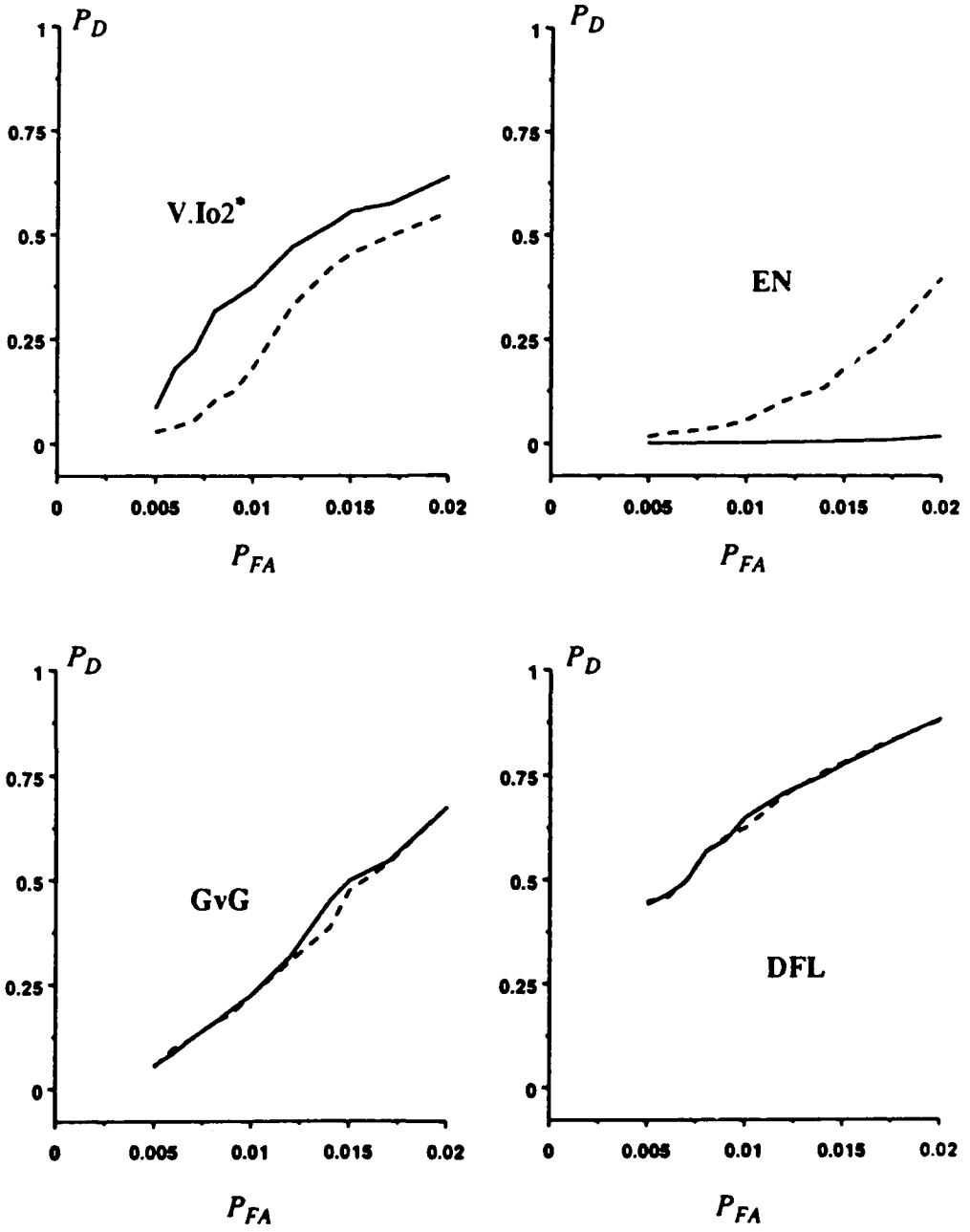

Fig. 6 — Filtered data, low-frequency

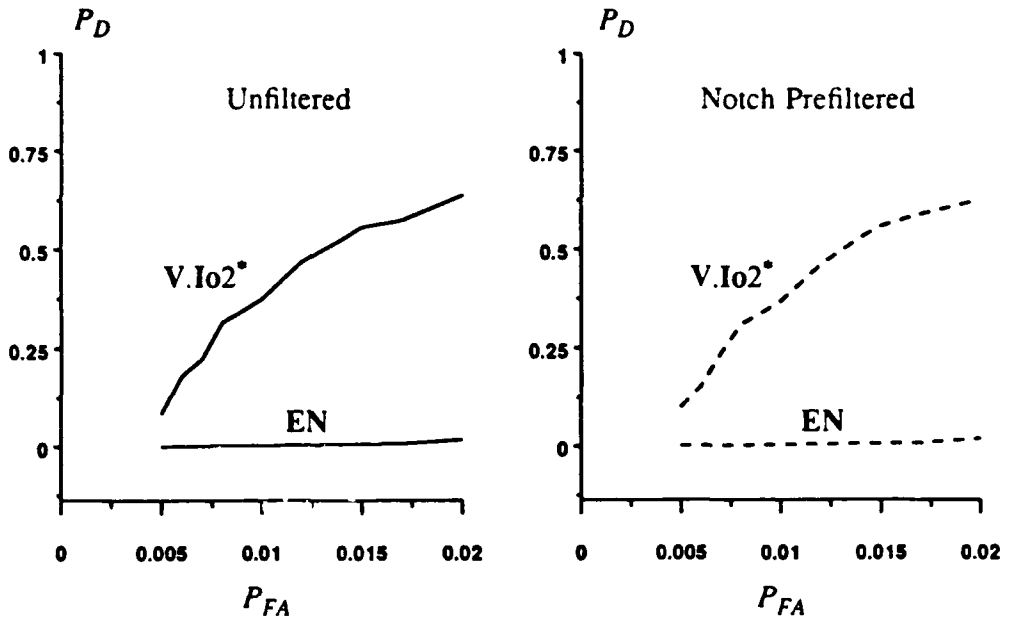

Fig. 7 — (U) Performance with notch filtering vs no filter

Figure 1: Simulated Data

Figure 1 shows performance for simulated data. For these evaluations, the same data set was used for training as for evaluations. The noise consisted of 5000 100-component vectors generated by the standard Wiener process, sampled at intervals of .01 second. The signal-plus-noise process consisted of 5000 100-component vectors generated by a diffusion with drift function f, f(x) = −25x. The standard Wiener process was used to generate the diffusion. The S + N process was then defined by

$$X([k+1]\Delta) = \Delta \sum_{i=1}^{k} (-25)X(i\Delta) + W([k+1]\Delta), \tag{4}$$

where W denotes the Wiener process. The sampled Wiener vectors used to generate the diffusion (X) vectors were not the same as those used to represent the noise.

The SNR (signal-to-noise ratio) is calculated according to

$$SNR = \frac{\text{Trace } \bar{R}_{S+N}}{\text{Trace } \bar{R}_N} - 1, \tag{5}$$

where $\bar{R}$ denotes correlation matrix (covariance matrix plus m*m, where m is the mean vector). This definition permits negative values of SNR. However, it gives the classical definition in the case of independent signal and noise. This seeming anomaly can be understood by noting that the classical SNR actually satisfies SNR + 1 = (S + N energy)/(noise energy), and this is also satisfied by the definition used here. Dependence between signal and noise can result in the S + N process having less energy than the N process.

For the simulated data, the SNR was -0.9, indicating strong negative correlation between the signal and noise components of the S + N process. According to theory, the V.IIo1 and GvG algorithms should have the same performance. In fact, the performance of GvG was slightly superior to that of V.II. This could be simply due to happenstance (finite data set). However, a more fundamental explanation may be more appropriate. Implementation of GvG depends only on knowledge of the data's covariance matrices and mean vectors (for N and S + N). Implementation of V.II requires knowledge of R_N, m_N, and the drift function. It may be that estimation of R_{S+N} is more accurate than estimation by regression of the drift, using 5000 data vectors. Also, errors in estimation of the drift may have more effect on V.II performance than errors in estimation of R_{S+N} have on performance of GvG.

Figure 1 also shows the performance of the adaptive V.Io2*. This detection algorithm has drift function σ modeled by

$$\sigma(x) = \sigma_0 + \sigma_1 x + \sigma_2 x^2, \tag{6}$$

when σ_0, σ_1, and $\hat{\sigma}_2 = \sigma_2/(1.6)$ are estimated by multiple linear regression.

Figure 1 also shows performance of V.Io1. V.Io1 has drift function σ given by

$$\sigma(x) = \sigma_0 + \sigma_1 x, \tag{7}$$

where σ_0 and σ_1 are estimated by linear regression. In principle, this should be the best-performing version of V.I, since the actual drift satisfies this model with $\sigma_0 = 0$ and $\sigma_1 = -25$. The superior performance of V.Io2* can be attributed to the relatively small number of data samples from which σ is estimated.

WEN had performance far inferior to that of the algorithms. Not displayed is the performance of the simple energy detector EN, which was even worse.

Figure 2: Large Training Ensembles

These curves show performance for the NOSC data using large (5000 vector) training ensembles of N and S + N data.

The curves include those for three algorithms whose implementation requires knowledge of S + N data properties: V.II, GvG, and DFL. THe V.IIo7* implementation has drift function σ which is seventh order polynomial:

$$\sigma_j(x) = \sum_{i=0}^{7} \sigma_{ji} x^i, \tag{8}$$

where the coefficients $\hat{\sigma}_{jk} = .625\ \sigma_{ji}$ for $i > 1$ and $\hat{\sigma}_{ji} = \sigma_{ji}$ for $i \leq 1$ were estimated by multiple linear regression.

Perhaps the most striking aspect of these results is the relatively poor performance of GvG. As previously discussed, the NOSC data were rejected by statistical testing as being Gaussian, although the tendency toward normality varied with the data set tested. However, there is no ambiguity about the detection results shown in Fig. 2: GvG performed far worse that V.Io7* and DFL. In fact GvG, which used a 5000-vector training ensemble from which to estimate its necessary S + N parameters, performed worse for P_{FA} values under .018 than the two implementations of V.Io2, which had only 100 data samples per observation from which to estimate S + N parameters.

Figure 2 also shows performance of algorithms that do not require prior knowledge of S + N properties. In keeping with the results on simulated data, V.Io2* (a) has drift function defined by

$$\sigma(x) = \sigma_0 + \sigma_1 x + \sigma_2 x, \tag{9}$$

where σ_0, σ_1, and $\hat{\sigma}_2$, $\hat{\sigma}_2 = \sigma_2/(1.6)$, are estimated by using multiple linear regression. However, weights other than 1.6 were also investigated; of these $\sigma_2 = \hat{\sigma}_2/.5$ gave the best results. This implementation is shown as V.Io2* (b).

Figure 3: Small Training Ensembles

These results are for training ensembles (N and S + N) of 200 sample vectors, as discussed in the preceding text. The same evaluation data were used as for the results of Fig. 2: 5000 from N and 5000 from S + N.

These results are considered more meaningful than those for the large training ensembles for most applications. The reason, of course, is that large training ensembles, particularly of the S + N process, will not usually be available. Even the noise characteristics may not be stable for long periods.

The most remarkable result of these evaluations is the superiority V.Io2* (1.6 factor, as in Fig. 1) at low values of P_{FA}. This is very promising for applications since V.I is easy to implement. It is particularly impressive in view of the very short observation time (100/4096 second) and the 100 sample values. Longer observation times and a larger number of data samples should improve the relative performance of V.I compared to GvG and DFL.

Results for V.II are not displayed. Those results were inferior to the V.I results using small training ensembles.

Figure 4: Time-averaged Drift, Version II

Version II permits time-varying σ in its implementation. Version I permits only time-invariant drift. With long observation times, the performance of Version I should be comparable to that of Version II using a time-averaged drift. That is, with the V.II original drift given by σ,

$$\sigma(j, x) = \sum_{i=0}^{p} \sigma(j, i) x^i, \tag{10}$$

define $\bar{\sigma}$ by

$$\bar{\sigma}(x) = \sum_{i=0}^{p} \bar{\sigma}_i x^i \tag{11}$$

where

$$\bar{\sigma}_i = \frac{1}{100} \sum_{j=1}^{100} \sigma_{ji}. \tag{12}$$

If long observation times are available, the estimate of a time-invariant σ should be reasonably close to the averaged time-varying σ.

Thus, the results of Fig. 4 give reasonable estimates of V.I performance if very long observation times are available. Of course, as elsewhere, this is relative performance. Longer observation times should also improve performance, if a good estimate of σ is available, without reference to how σ was obtained.

The best results were obtained by averaging V.IIo2*, which was original drift σ given by

$$\sigma_k(x) = \sigma_{k0} + \sigma_{k1}x + \sigma_{k2}x^2, \tag{13}$$

with σ_{k0}, σ_{k1}, and $\hat{\sigma}_{k2} = \sigma_{k2}/(1.6)$ determined from multiple linear regression. The remarkable aspect of these results can be seen by comparing them with those given for V.IIo7* in Fig. 2. This reveals that the time-averaged second-order modified drift gave performance quite comparable to that with the time-varying seventh order drift. Extrapolating, one may speculate that performance of V.Io2* may show the same superiority over GvG and DFL as given by V.IIo7* when long observation times are available. The hypothesis needs to be investigated; if adequate experimental data are not available then simulations could be used for a partial evaluation.

Figure 5: Version II Performance, Various Orders of Polynomial Drift

Figure 5 shows the performance of the V.II algorithms using unmodified polynomial regression to estimate the drift, with orders ranging from one to eight. Improvement over Order 1 occured almost entirely at P_{FA} values below .009. More evaluations are necessary to determine if the improvement justifies the complexity. Compared to V.IIo7*, shown in Fig. 2, the difference is more significant, but the relative superiority of V.IIo7* was still not impressive at P_{FA} values greater than .009.

An interesting aspect of these results is the performance of V.IIo1, which uses a first-order polynomial, in comparison with the performance of GvG as shown in Fig. 2. As previously discussed, if S + N and N are both Gaussian, then the performance of V.IIo1 and GvG should be the same. Recall that GvG had performance slightly better than that of V.IIo1 for the Gaussian simulated data (Fig. 1). However, for the sonar data, V.IIo1 far outperformed GvG. This is another indication of the significant effect of the non-Gaussian property of the data.

Figure 6: Filtered Data, Low-Frequency

Low-frequency data were obtained by passing the original data through a low-frequency bandpass filter and also through two notch filters. The notch filters were inserted to remove weak lines in the spectra.

Figure 6 shows performance using this filtered data with large training ensembles (5000 vectors for N and for S + N). The data sets were those used to construct Fig. 2, but after being passed through the bandpass filter and the two notch filters. For comparison, the results for the unfiltered data, already seen in Fig. 2, are repeated.

The SNR for this filtered data was 2.886, vs 1.885 for the unfiltered data. The improved SNR should result in all detection algorithms improving their performance. However, if an algorithm is initially optimum for unfiltered data, its performance cannot improve for filtered data, since the filtering simply introduces another stage into the detection algorithm. Conversely, a suboptimum detection algorithm may well have improved performance if preceded by narrowband filtering.

These general considerations are borne out by the results shown in Fig. 6. V.Io2* is implemented exactly as for Fig. 2. The results from Fig. 2 are shown for comparison. The algorithm's performance on the filtered data substantially decreased. By contrast, the performance of the simple energy detector, EN, substantially improved.

The basic idea here is that bandpass filtering is an irreversible operation on the data. Thus, information available to an optimum algorithm is lost when bandpass filtering is applied, unless the filtering is equivalent to a stage in the operation of the optimum algorithm on the unfiltered data. Evidently, for the V.Io2* algorithm this is not the case. This can be understood by noting that the bandpass filtering not only removes noise but also removes signal components in the higher frequency ranges.

Figure 6 also shows the performances of GvG and DFL for filtered and unfiltered data. As can be seen, there is very little difference between the results for the filtered and unfiltered data for these two algorithms.

One complicating factor here is that the filtering included notches as well as the bandpass. One could speculate that the degradation in performance of V.Io2* is due to loss of the line components. That this is not the case seen in Fig. 7.

Figure 7: Performance with Notch Filtering vs No Filter

Figure 7 shows performance of V.Io2* and EN. Performance is shown for unfiltered data and for filtered data when the filter consists only of notches.

There was very little difference between performances of the two algorithms on filtered and unfiltered data. For V.Io2*, this shows that the degradation due to filtering shown in Fig. 6 was due to the low-frequency bandpass filtering.

CONCLUSIONS AND RECOMMENDATIONS

The study reported here compared the detection performance of two new algorithms with a number of appropriate reference algorithms using both simulated data and passive sonar data. One of the more interesting results of the study was the relatively poor performance with the sonar data of the reference algorithm that is a log-likelihood ratio under the assumption that both N and S + N are Gaussian. Repetition of such a result using more extensive data would be an important and perhaps largely unexpected revelation regarding models for passive sonar. The fact that sonar data fails a statistical test for normality may not necessarily imply that algorithms based on the assumption of normality will not perform well. However, the results given here indicate that the nature of the departure from Gaussian for this data set was serious from a signal detection viewpoint, thus providing more incentive for development of algorithms that are not based on the assumption of Gaussian data.

The results summarized here are the first computational results obtained for the new algorithms. The broadband character of the sonar signal data is an important aspect of the work. As previously discussed, the algorithms are directly descended, via reasonable assumptions and appropriate approximations, from the exact log-likelihood-ratio for the continuous-time data.[3] Although their optimality is based on the assumption of Gaussian noise, they require no assumptions about the statistical properties of the S + N process. Moreover, the results of the present study, in which the sonar noise data failed statistical tests for normality, indicate that the algorithms may not be sensitive to modest departures from normality of the noise process. Their performance on the passive sonar data used in this study was superior to that of all the comparable reference algorithms, despite the very rudimentary method used to estimate the unknown drift function. The performance of both algorithms should improve when this procedure is optimized. In addition, the extremely short observation time and small number of data samples per observation, necessary because of data limitations, were presumably a strong handicap to the adaptive version. The fact that the algorithms performed so well under these handicaps is very encouraging, although we emphasize that this is for only one data set. Nevertheless, the results provide a preliminary confirmation of the theoretical advantages of the new algorithms, particularly for use in detecting broadband signals. Extensive further work, both theoretical and computational, is now needed to realize their full potential.

Some of this work is fairly evident; other parts less so. It includes optimization of the method for estimating the diffusion drift function, development of optimum array processing (for both fixed and movable arrays) based on the new single-time-series algorithms, investigation of the effect of sampling rate on the performance of approximations to likelihood ratios derived from S + N that is a filtered diffusion (and particularly for the adaptive Version I), development of efficient and reliable methods of simulating filtered diffusions (needed because large ensembles of S + N data are likely to be unavailable for many applications, while empirical evaluations will be needed to determine performance estimates), and extension of the detection algorithms to classification.

The theoretical components of the additional work required can be anticipated to be rather complex. The reason is that the algorithms are not obtained by considering any optimality criterion for the discrete-time problem but as approximations to a continuous-time likelihood ratio. Thus, it is the continuous-time problem that is at the heart of the algorithms's development, as described in Refs. 1 and 3. This point needs to be constantly kept in mind in carrying out the further work described above. That work seems clearly worthwhile: The new algorithms appear to have the potential of providing significant improvements over existing detection and classification methods when the signals are broadband, without any assumptions on the signals' statistical properties.

ACKNOWLEDGMENTS

The work reported here was done at the Laboratory for Information and Stochastic Systems, University of North Carolina, and was supported by ONR Grant N00014-J-89-1175. The authors thank R. Lund and P.L. Seymour for valuable help with programming. Thanks are also extended to the referee and to the editor for their comments.

REFERENCES

1. C.R. Baker and A.F. Gualtierotti, "Discrimination with Respect to a Gaussian Process," Probabil. Theory and Related Fields **71**, 159-182 (1986).

2. C.R. Baker and A.F. Gualtierotti, "Likelihood Ratios and Signal Detection for NonGaussian Process," *Stochastic Processes in Underwater Acoustics*, C.R. Baker, ed., Lecture Notes in Control and Information Science **85**, 154-180 (Springer-Verlag, Berlin, 1986).

3. C.R. Baker and A.F. Gualtierotti, "Likelihood-Ratio Detection of Stochastic Signals," LISS 40, Department of Statistics, University of North Carolina at Chapel Hill (Oct., 1989).

4. R.S. Lipster and A.N. Shiryayev, *Statistics of Random Processes I. General Theory* (Springer-Verlag, New York, 1977).

5. L.W. Weinstein and V.D. Zubakov, *Extraction of Signals from Noise* (Prentice-Hall, London, 1962).

6. C.R. Baker, "Optimum Quadratic Detection of a Random Vector in Gaussian Noise," IEEE Trans. Commun. Tech. **14**, 802-805 (1966).

7. C.R. Baker, "On the Deflection of a Quadratic-Linear Test Statistic," IEEE Trans. Inf. Theory **15**, 16-21 (1969).

8. R.M. Bethes, B.S. Duran, and T.L. Boullion, *Statistical Methods for Engineers and Scientists* (Marcel Dekker, New York, 1975).

9. C.R. Baker, "Statistical Tests for the Analysis of Sonar Data," JUA(USN) **26**, 227-245 (1976).